working paper
department
of economics

REGRESSION-BASED INFERENCE IN LINEAR TIME SERIES MODELS WITH INCOMPLETE DYNAMICS

Jeffrey M. Wooldridge

No. 550 April 1990

massachusetts
institute of
technology
50 memorial drive
cambridge, mass. 02139

REGRESSION-BASED INFERENCE IN LINEAR TIME SERIES MODELS WITH INCOMPLETE DYNAMICS

Jeffrey M. Wooldridge

No. 550 April 1990

REGRESSION-BASED INFERENCE IN LINEAR TIME SERIES MODELS WITH INCOMPLETE DYNAMICS

Jeffrey M. Wooldridge
Department of Economics, E52-262C
Massachusetts Institute of Technology
Cambridge, MA 02139
(617) 253-3488

August 1989
Latest Revision: April 1990

Abstract

Regression-based heteroskedasticity and serial correlation robust standard errors and specification tests are proposed for linear models that may not represent an expectation conditional on all past information. The statistics are computable via a sequence of linear regressions, and the procedures apply to models estimated by ordinary least squares or two stage least squares. Examples of the specification tests include tests for nonlinearities in static models, exclusion restriction tests in finite distributed lag models, heteroskedasticity/serial correlation-robust Chow tests, tests for endogeneity, and tests of overidentifying restrictions. Some new tests of the assumptions underlying Cochrane-Orcutt estimation are also proposed, and some considerations when applying the various robust tests are discussed.

1. Introduction

Work by Hansen (1982), Domowitz and White (1982), White (1984) and, more recently, by Newey and West (1987), Gallant (1988), and Gallant and White (1988), has provided general methods for performing inference in econometric models that may be dynamically incomplete. A simple example of a dynamically incomplete model is a static regression model with neglected serial correlation. As discussed by Hansen and Hodrick (1980) and Hansen (1982), more complicated examples arise in rational expectations models when the time interval relevant for decision making by economic agents differs from the sampling interval. In these models the implied errors are not martingale difference sequences but moving averages of a particular order. The generalized method of moments procedures unified by Hansen (1982) are used regularly in these rational expectations-type applications.

Serial correlation robust procedures have been much slower to catch on for standard linear regression applications. There are probably several reasons for this, the foremost being the availability of a competitor that is implemented in all regression packages, namely the simple AR(1) model that can be estimated by the Cochrane-Orcutt technique. Estimating a static or finite distributed lag model with an AR(1) correction is now as easy as estimating the model by OLS. On the other hand, heteroskedasticity/serial correlation-robust (H/SC-robust) covariance matrix estimators suggested by Domowitz and White (1982) and Newey and West (1987) are more difficult to compute, and many regression packages do not report them. Currently available H/SC-robust specification tests of more than one degree-of freedom introduce even more complications. Specification testing in the AR(1) model is straightforward because all statistics can be computed as standard F-tests based on quasi-differenced data.

Certain limitations of the AR(1) model have been stressed by many authors; for a summary and references see Hendry, Pagan, and Sargan (1984) (hereafter HPS (1984)). As emphasized by these authors, one way to view the static model with AR(1) errors is as a restricted version of a general dynamic model. The restrictions on the dynamic model have become known as "common factor" restrictions, and various tests of these restrictions are available (e.g. HPS (1984) and Harvey (1981)). If the common factor restrictions are violated, then Cochrane-Orcutt type estimators may be inconsistent for the parameters appearing the in static relationship. At a minimum, one would have to question the validity of the usual OLS test statistics based on quasi-differenced data.

By focusing on the common factor restrictions one necessarily adopts the viewpoint that the dynamic expectation is of primary importance. As argued in section 2, the validity of the common factor restrictions in the dynamic regression is neither necessary nor sufficient for methods such as Cochrance-Orcutt (C-O) or nonlinear least squares (NLS) to consistently estimate the parameters in the static relationship. Consequently, other testing procedures are needed to assess whether C-O estimates are consistent for the parameters of the static relationship. Some new tests useful in this regard are offered in in section 4.

An alternative to the static plus AR(1) model is to correct the OLS standard errors for serial correlation, and to compute H/SC-robust specification tests. The primary purpose of this paper is to offer forms of these statistics that can be computed by virtually any regression package. Thus, these techniques are only modestly more costly than Cochrane-Orcutt in terms of computation, while being more robust and more widely applicable.

Section 2 of the paper reviews limiting distribution results for a linear model with heteroskedasticity and/or serial correlation of unknown form. The computationally simple H/SC-robust standard errors suggested by Wooldridge (1989) are presented and extended. The static/AR(1) model is also presented under the assumptions most useful for the current analysis. Section 3 develops regression-based H/SC-robust specification tests, along with several examples. Some considerations when applying these tests are discussed in section 4, along with some heteroskedasticity-robust tests for common factor restrictions. Section 5 contains the methods appropriate for two stage least squares estimation, and section 6 contains some suggestions for future research.

2. Background and Motivation

Let $\{(y_t,z_t): t=\ldots-1,0,1,,\ldots\}$ be a strictly stationary, ergodic time series, where y_t is a scalar and z_t is a 1xJ vector. Due to the work of White (1984) and others, it is well-known that strict stationarity can be relaxed by imposing mixing or other weak dependence requirements. However, the dependence and moment conditions imposed typically rule out integrated processes or series with deterministic trends; therefore, there are no practical consequences of the strict stationarity assumption when analyzing correctly specified models with weakly dependent series that have some bounded moments. Unit root processes are ruled out in what follows because some of the specification tests would have nonstandard limiting distributions. Although not treated explicitly, series with deterministic polynomial trends are easily handled in the usual manner by including an appropriate polynomial trend in the estimation (so that the data are appropriately detrended). In all of the subsequent calculations (particularly auxiliary regressions), the functions of time can be used in the same manner as the stationary regressors.

In a time series context, there are several relationships between y_t and z_t that one might be interested in. The simplest model relating economic time series is the *static model*, which focuses on the contemporaneous relationship between y_t and z_t, ignoring any dynamic aspects. In particular, interest centers on the conditional expectation $E(y_t|z_t)$, so that the static model is similar in spirit to cross section regression models. However, due to the dependent nature of the data, the errors in static time series regression models display serial correlation.

Assuming linearity of the conditional expectation, the static model can be written as

(2.1) $$E(y_t|z_t) = \alpha + z_t\delta, \qquad t=1,2,...$$

or, in error form,

(2.2) $$y_t = \alpha + z_t\delta + u_t, \quad E(u_t|z_t) = 0, \qquad t=1,2,....$$

Estimating a model for $E(y_t|z_t)$ is reasonable if one is interested in the contemporaneous effect of z_t on y_t. The researcher must decide if the conditional expectation (2.1) is of interest. Sometimes y and z are more properly viewed as being jointly determined, in which case δ is still well-defined but not of much interest.

Except for standard regularity conditions, $E(u_t|z_t) = 0$ is sufficient for the ordinary least squares estimator to be consistent for α and δ. In particular, there is no need to assume

(2.3) $$E(u_t|...,z_{t+1},z_t,z_{t-1},...) = 0,$$

which is a strict exogeneity assumption on $\{z_t\}$ and operationally the same as assuming nonrandomness of $\{z_t\}$. Further, even when the errors $\{u_t\}$ contain substantial serial correlation OLS estimates are generally consistent and asymptotically normally distributed.

Another relationship frequently of interest to economists is the distributed lag of y on z. This relationship allows one to trace the pattern of the dynamic effects of a change in contemporaneous z on subsequent values of y. The expecation of interest is the expectation of y_t given the current and past values of z, $E(y_t|z_t,z_{t-1},...)$. If it is assumed that the effect of z_{t-j} on y_t is zero for $j > Q$ then

$$E(y_t|z_t,z_{t-1},...) = \alpha + z_t\delta_0 + z_{t-1}\delta_1 + ... + z_{t-Q}\delta_Q \tag{2.4}$$

or

$$y_t = \alpha + z_t\delta_0 + z_{t-1}\delta_1 + ... + z_{t-Q}\delta_Q + u_t, \quad E(u_t|z_t,z_{t-1},...) = 0. \tag{2.5}$$

From a statistical viewpoint the error assumption in (2.5) could be replaced by the weaker requirement

$$E(u_t|z_t,...,z_{t-Q}) = 0, \tag{2.6}$$

but then the δ_j would be more difficult to interpret. In what follows a finite distributed lag model is defined by (2.4) or (2.5).

As with the static model, the strict exogeneity assumption (2.3) is not required for OLS to consistently estimate δ_0, δ_1, ..., and δ_Q. Also, the errors $\{u_t\}$ in (2.5) will generally exhibit serial correlation and heteroskedasticity.

The static and finite distributed lag models are special cases of the statistical model

$$(2.7) \qquad y_t = x_t\beta + u_t, \quad E(u_t|x_t) = 0, \quad t=1,2,\ldots,$$

where x_t is a 1xK subvector of $(1,z_t,y_{t-1},z_{t-1},\ldots)$ (the lag lengths appearing in x_t are necessarily invariant across t). Before proceeding to the statistical analysis of (2.7), it is important to stress that, of the models discussed so far -- the static, distributed lag, and general dynamic models (i.e. x_t contains lags of y as well as lags of z) -- none is necessarily the "true" model or the "true" data generating mechanism. The models simply correspond to different conditional expectations of the variable y_t. Ideally, economic theory specifies whether $E(y_t|z_t)$, $E(y_t|z_t,z_{t-1},\ldots)$, $E(y_t|y_{t-1},z_{t-1},\ldots)$, or some other expectation is the one of interest. For example, rational expectations places restrictions on expectations of the form $E(y_t|y_{t-j},z_{t-j},y_{t-j-1},z_{t-j-1},\ldots)$ for some integer $j \geq 1$. But much of the time it is up to the researcher to specify which relationship is of interest.

In the general model (2.7), the law of iterated expectations implies that x_t and u_t are uncorrelated: $E(x_t'u_t) = E[E(x_t'u_t|x_t)] = E[x_t'E(u_t|x_t)] = 0$. Importantly, this is true whether or not x_t contains lagged dependent variables and whether or not $\{u_t\}$ is serially correlated. Provided that $E(y_t|x_t)$ is of interest, OLS consistently estimates the coefficients appearing in this expectation under general regularity conditions. Because these conditions are covered in detail by White (1984), as a starting point it is assumed that the OLS estimator

$$(2.8) \qquad \hat{\beta} = (X'X)^{-1}X'Y = \beta + \left(T^{-1}\sum_{t=1}^{T} x_t'x_t\right)^{-1} T^{-1}\sum_{t=1}^{T} x_t'u_t$$

is asymptotically normally distributed. More precisely,

$$(2.9) \qquad \sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, A^{-1}BA^{-1}),$$

where

$$A \equiv E(X'X/T) = E(x_t'x_t),$$

$$B \equiv \lim_{T\to\infty} V\left[T^{-1/2}\sum_{t=1}^{T} x_t'u_t\right] = \Omega_0 + \sum_{j=1}^{\infty} [\Omega_j + \Omega_j'],$$

and

$$\Omega_j \equiv E(s_{t+j}s_t'), \quad s_t \equiv x_t'u_t.$$

The structure of B reveals that, although the serial correlation structure of $\{s_t \equiv x_t'u_t\}$ does not affect the ability of OLS to estimate $E(y_t|x_t)$, it does manifest itself in the limiting distribution of the OLS estimator. Nevertheless, provided one has consistent estimators $\hat{A}$ of A and $\hat{B}$ of B, in practice one carries out inference on β as if

$$\text{"}\hat{\beta} \sim N(\beta, \hat{A}^{-1}\hat{B}\hat{A}^{-1}/T)\text{."}$$

$\hat{A}^{-1}\hat{B}\hat{A}^{-1}/T$ is an estimator of the asymptotic variance of $\hat{\beta}$, $A^{-1}BA^{-1}/T$. A consistent estimator of A in the present context is simply $\hat{A} \equiv X'X/T$. Estimation of B is generally more difficult because, as seen above, B generally depends on the autocorrelation and variance structure of $\{x_t'u_t: t=1,2,\ldots\}$. Before proceeding to the general case it is useful to recall the conditions that provide asymptotic justification for the use of the usual t-statistics and F-statistics.

The appropriate no serial correlation assumption with possibly random regressors is

(2.10) $$E(u_{t+j}u_t|x_{t+j},x_t) = 0, \; j \geq 1.$$

Although (2.10) implies that the $\{u_t\}$ are unconditionally uncorrelated, i.e. $E(u_{t+j}u_t) = 0$, $j \geq 1$, this latter condition is generally insufficient for the usual test statistics to be approximately valid (unless the x_t are treated as nonrandom). The important consequence of (2.10) is

$$\Omega_j = E(x'_{t+j}u_{t+j}u_t x_t) = 0,$$

which follows from (2.10) by a straightforward application of the law of iterated expectations. Consequently, under (2.10) B reduces to the simple formula

$$B = T^{-1}\sum_{t=1}^{T} E(u_t^2 x_t' x_t) = E(u_t^2 x_t' x_t).$$

The appropriate homoskedasticity assumption with possibly random regressors is

$$E(u_t^2 | x_t) = \sigma^2. \quad (2.11)$$

As with serial correlation, (2.11) imposes <u>conditional</u> homoskedasticity on u_t (or y_t); it is not enough that $E(u_t^2)$ be constant across t (which is always the case here by stationarity). If x_t contains lagged dependent variables then (2.11) rules out certain dynamic forms of heteroskedasticity, such as Engle's (1982) ARCH model.

If (2.11) holds in addition to (2.10) then another application of the law of iterated expectations gives $B = \sigma^2 E(x_t' x_t) = \sigma^2 A$, and the asymptotic variance of $\sqrt{T}(\hat{\beta} - \beta)$ reduces to the well-known formula

$$AV\ \sqrt{T}(\hat{\beta}_T - \beta) = \sigma^2 [E(X'X/T)]^{-1}.$$

Estimating σ^2 by the usual degrees-of-freedom adjusted estimator

$$\hat{\sigma}^2 = (T-K)^{-1}\sum_{t=1}^{T}(y_t - x_t\hat{\beta}_T)^2 = SSR/(T-K) \quad (2.12)$$

produces the usual standard errors and test statistics, which are asymptotically valid under (2.7), (2.10) and (2.11).

The homoskedasticity assumption (2.11) is a convenience that can never be guaranteed to hold a priori: a model for $E(y_t|x_t)$ by definition imposes no restrictions on $V(y_t|x_t)$. In contrast, there is one well-known case where the no serial correlation assumption is satisfied. Suppose that u_t is unpredictable given x_t and all past information $\phi_{t-1} \equiv (y_{t-1},x_{t-1},y_{t-2},x_{t-2},\ldots)$ (equivalently, ϕ_{t-1} can be taken to be $(u_{t-1},x_{t-1},u_{t-2},x_{t-2},\ldots)$). More formally, the condition is

$$E(u_t|x_t,\phi_{t-1}) = 0, \qquad (2.13)$$

which implies that

$$E(y_t|x_t,\phi_{t-1}) = E(y_t|x_t) = x_t\beta. \qquad (2.14)$$

When (2.14) holds, x_t contains enough lags of y and/or z (which in principle could be no lags of y or z) so that additional lags do not help to predict y_t; if this is the case, model (2.7) is said to be *dynamically complete*. Dynamic completeness is easily seen to imply the no serial correlation assumption (2.10): because $(x_{t+j},u_t,x_t) \subset (x_{t+j},\phi_{t+j-1})$, it follows that $E(u_{t+j}|x_t,\phi_{t+j-1}) = 0$; the law of iterated expectations then implies

$$E(u_{t+j}|x_{t+j},u_t,x_t) = 0, \quad \text{all } j \geq 1, \qquad (2.15)$$

which implies (2.10) and $B = E(u_t^2 x_t' x_t)$. If heteroskedasticity is present, the usual covariance matrix estimator must be modified. The White (1980) heteroskedasticity-robust covariance matrix estimator is easily shown to be consistent in time series applications with no serial correlation; see also Hsieh (1983).

If the model is not dynamically complete then (2.10) typically fails and the usual and White covariance matrix estimators are inconsistent. For static (and finite distributed lag) models, a popular procedure when serial correlation is detected is to assume that the errors follow an AR(1) process. As it is usually analyzed, this model can be expressed as

(2.16) $$y_t = \alpha + z_t\delta + u_t, \quad E(u_t|z_t) = 0, \quad t=1,2,\ldots.$$

(2.17) $$u_t = \rho u_{t-1} + e_t, \quad E(e_t|z_t,\phi_{t-1}) = 0, \quad t=1,2,\ldots, \quad |\rho| < 1,$$

with the additional, although less important, homoskedasticity assumption $E(e_t^2|z_t,\phi_{t-1}) = \sigma_e^2$. In section 4 it will be shown that these assumptions on e_t are in fact stronger than needed for the Cochrane-Orcutt method to consistently estimate $\beta \equiv (\alpha,\delta')'$. However, assumptions (2.16) and (2.17) (and $E(e_t^2|z_t,\phi_{t-1}) = \sigma_e^2$) ensure that the usual statistics based on quasi-differenced data (with the estimated ρ) are valid. Also, these are the assumptions underlying the usual LM test for H_0: $\rho = 0$. The homoskedasticity assumption is less important because it can be shown that using heteroskedasticity-corrected test statistics in the quasi-differenced regressions produces valid test statistics. (This follows because, letting $\tilde{\beta}$ and $\tilde{\rho}$ denote the C-O or NLS estimators of $\beta \equiv (\alpha,\delta')'$ and ρ, (2.16) and (2.17) ensure that the limiting distribution of $\sqrt{T}(\tilde{\beta} - \beta)$ does not depend on that of $\sqrt{T}(\tilde{\rho} - \rho)$.)

It is important to observe that (2.16) is a model of the static conditional expectation $E(y_t|z_t)$; (2.17) then implies a particular form for the dynamic conditional expectation:

(2.18) $$E(y_t|z_t,\phi_{t-1}) = (1-\rho)\alpha + z_t\delta + \rho(y_{t-1} - z_{t-1}\delta), \quad t=1,2,\ldots.$$

Letting $x_t \equiv (1,z_t,y_{t-1},z_{t-1})$ shows that (2.18) can be expressed as a dynamic linear model such that $E(y_t|x_t) = E(y_t|x_t,\phi_{t-1})$, i.e. (2.18) is dynamically complete. Also, note that if (2.16) and (2.17) hold, and $\rho \neq 0$, then (by the law of iterated expectations) it must be the case that

(2.19) $$E(u_t | z_{t+1}) = 0;$$

this is a type of exogeneity condition on the explanatory variables that is imposed by the AR(1) model. In section 4, this condition is shown to be critical for C-O to produce consistent estimates of δ.

An unrestricted version of (2.18) is

(2.20) $$y_t = \alpha_0 + z_t\delta_0 + \rho_1 y_{t-1} + z_{t-1}\delta_1 + e_t,$$

and the *common factor* restrictions are embodied in the J nonlinear constraints

(2.21) $$\delta_1 = -\rho_1\delta_0.$$

Under (2.16) and (2.17), the regression

$$y_t \text{ on } 1, z_t, y_{t-1}, z_{t-1}$$

consistently estimates $\delta = \delta_0$ as the coefficient vector on z_t. However, it is important to see that (2.20) and (2.21) can hold with δ_0 bearing no resemblance to δ. As an example, suppose that $\delta_0 = \delta_1 = 0$ in (2.20), so that (2.21) is trivially true, while δ in (2.16) is different from zero. Nothing about (2.16), (2.20), and (2.21) rules out this possibility. On the other hand, a rejection of (2.21) in the context of model (2.20) tells one nothing about the static relationship; $E(y_t|z_t)$ is still well-defined and potentially of interest.

In fact, posing the AR(1) model as (2.16) and (2.17) suggests that at least some interest lies in the vector δ describing the contemporaneous relationship between y_t and z_t. The AR(1) assumption justifies the use of the Cochrane-Orcutt method to obtain standard errors, t-statistics, and other test statistics that have the usual interpretations and are asymptotically optimal (assuming that $E(e_t^2|z_t,\phi_{t-1})$ is constant). In most of the common factor literature -- e.g. Sargan (1964,1980), Hendry and Mizon (1978), Hendry, Pagan, and Sargan (1984) -- the existence of common factors such as those implied by the AR(1) model is interpreted as implying that the relationship between y_t and z_t is static, with the dynamics entering only through the error term. On the other hand, rejection of the common factor restrictions is viewed as implying a dynamic relationship between y and z, and therefore the unrestricted model (2.20) should be estimated. A different perspective is that the vector δ is well-defined whether or not the common factor restrictions hold; it is simply the case that $\delta \neq \delta_0$, so that the link between the static and dynamic expectations has been broken. Without a specific context it is unclear whether δ is of less interest simply because a certain set of nonlinear constraints on the parameters of the dynamic expectation are not satisfied.

An important consequence of the preceding discussion is that the conditions underlying the consistency and, more generally, the validity of the usual test statistics in the the static (or DL) model with AR(1) errors are not innocuous. If interest lies in the static relationship $E(y_t|z_t)$, the DL relationship $E(y_t|z_t,z_{t-1},\ldots)$, or some other expectation that is dynamically incomplete, without also imposing assumptions on the fully dynamic conditional expectation $E(y_t|z_t,\phi_{t-1})$, or on the exogeneity properties of z_t, then it is possible to compute serial correlation robust standard errors of the OLS estimator. The no serial correlation assumption (2.14) can be replaced by an assumption ensuring that the dependence in $\{x_t'u_t\}$ dies off sufficiently fast for B to be consistently estimated. For robustness reasons estimating $E(y_t|z_t)$ via a static regression and computing corrected standard errors is often preferred to performing an AR(1) correction. Testing the common factor restrictions is reviewed in section 4.

A heteroskedasticity/serial correlation-robust estimator of B has been recently proposed by Newey and West (1987) and Gallant and White (1988); both papers modify an estimator due to White and Domowitz (1984). The estimator is given by

$$(2.22) \qquad \hat{B} \equiv \hat{\Omega}_0 + \sum_{j=1}^{G} \varphi(j,G)[\hat{\Omega}_j + \hat{\Omega}_j'],$$

where

$$\hat{\Omega}_j \equiv (T-K)^{-1} \sum_{t=j+1}^{T} \hat{s}_t' \hat{s}_{t-j}, \quad \hat{s}_t \equiv x_t \hat{u}_t$$

and

$$(2.23) \qquad \varphi(j,G) \equiv 1 - j/(G+1), \quad j=1,\ldots,G$$
$$\equiv 0, \quad j=G+1,G+2,\ldots$$

are weights that have been used in the literature on spectral density estimation, and G is a nonnegative integer. As Newey and West (1987) show, the weighting in (2.22) ensures that $\hat{B}$ is positive semi-definite. The degrees of freedom adjustment factor $(T-K)^{-1}$ has been used in the definition of $\hat{\Omega}_j$ because there is some evidence that it reduces finite sample bias. Given the estimator $\hat{B}$, the heteroskedasticity/serial correlation-robust covariance matrix estimator of $\hat{\beta}$ is given by

$$(2.24) \qquad \hat{V}/T \equiv (X'X/T)^{-1}\hat{B}(X'X/T)^{-1}/T = (X'X)^{-1}(T\hat{B})(X'X)^{-1}.$$

The asymptotic standard error of $\hat{\beta}_j$ is obtained as the square root of the jth diagonal element of this matrix.

Sometimes it is useful to be able to avoid the matrix manipulations involved in computing (2.24). By focusing on one variance (or covariance) at a time, an H/SC-robust estimator can be obtained from simple OLS regressions. Wooldridge (1990b) shows that the following procedure is valid for computing an H/SC-robust standard error for $\hat{\beta}_j$.

PROCEDURE 2.1:

(i) Run the regression

$$y_t \text{ on } x_{t1}, x_{t2}, \ldots, x_{tK}, \quad t=1,\ldots,T \tag{2.25}$$

and obtain "se($\hat{\beta}_j$)", $\hat{\sigma}$, and the residuals $\{\hat{u}_t : t=1,\ldots,T\}$. Here "se($\hat{\beta}_j$)" denotes the usual (generally incorrect) standard error reported for $\hat{\beta}_j$, and $\hat{\sigma}$ is the standard error of regression (2.25).

(ii) Run the regression

$$x_{tj} \text{ on } x_{t1},\ldots,x_{t,j-1},x_{t,j+1},\ldots,x_{tK}, \quad t=1,\ldots,T \tag{2.26}$$

and save the residuals, say $\{\hat{r}_{tj} : t=1,\ldots,T\}$.

(iii) Define $\hat{\xi}_t \equiv \hat{r}_{tj}\hat{u}_t$ and let

$$\hat{c}_j \equiv \{\hat{\omega}_0 + 2\sum_{s=1}^{G} \varphi(s,G)\hat{\omega}_s\} \tag{2.27}$$

where

$$\hat{\omega}_s \equiv (T-K)^{-1} \sum_{t=s+1}^{T} \hat{\xi}_t\hat{\xi}_{t-s}, \quad s = 0,\ldots,G,$$

$\varphi(s,G)$ is given by (2.23), and G is, say, the integer part of $T^{1/4}$.

Alternatively, compute $\hat{c}_j$ as

$$\hat{c}_j \equiv [T/(T-K)]\hat{\tau}_G^2/(1 - \hat{a}_1 - \hat{a}_2 - \ldots - \hat{a}_G)^2, \tag{2.28}$$

where $\hat{a}_i$, $i=1,\ldots,G$, are the OLS coefficients from the autoregression

$$\hat{\xi}_t \text{ on } \hat{\xi}_{t-1},\ldots,\hat{\xi}_{t-G}, \tag{2.29}$$

and $\hat{\tau}_G^2$ is the square of the usual standard error of regression (2.29).

(iv) The H/SC-robust standard error of $\hat{\beta}_j$ is

$$(2.30) \qquad se(\hat{\beta}_j) = \left[\sum_{t=1}^{T} \hat{r}_{tj}^2\right]^{-1} (T\hat{c}_j)^{1/2} = ["se(\hat{\beta}_j)"/\hat{\sigma}]^2 (T\hat{c}_j)^{1/2}. \quad ■$$

Equation (2.30) offers a simple adjustment to the usual OLS standard error that is robust to heteroskedasticity and serial correlation, which simply requires the additional OLS regressions (2.26) and (2.29).

Note that $\hat{c}_j$ given by (2.27) is simply (2.22) applied to the scalar sequence $\{\hat{\xi}_t = \hat{r}_{tj}\hat{u}_t: t=1,\ldots,T\}$; it is a consistent estimator of the spectral density of $\{\xi_t\}$ at frequency zero. The estimator (2.28) is Berk's (1974) autoregressive spectral density estimator. Berk requires $G = o(T^{1/4})$ to establish consistency. If the model is dynamically complete, so that there is no serial correlation present, then a heteroskedasticity-robust standard error is obtained from Procedure 2.1 by setting $\hat{c}_j = (T-K)^{-1} \sum_{t=1}^{T} \hat{\xi}_t^2$.

The H/SC-consistent standard errors allow construction of t-statistics for testing individual hypotheses about $\beta_1,\ldots,\beta_K$. Note that the choice of G can be different for each $\hat{\beta}_j$; it is the serial correlation properties of $\{r_{tj}u_t\}$ that matter for the standard error of $\hat{\beta}_j$. Because the standard error of any linear combination of β can be obtained via an OLS regression on transformed variables, robust standard errors of linear combinations are easily computed using Procedure 2.1. For example, a robust standard error for the long run propensity in a distributed lag model can easily be computed. A robust estimator for the covariance between any two coefficients, say $\hat{\beta}_i$ and $\hat{\beta}_j$, is easily obtained from $V(\hat{\beta}_i)$, $V(\hat{\beta}_j)$, and, say, $V(\hat{\beta}_i + \hat{\beta}_j)$. Simply use the asymptotic analog of the relationship

$$CV(\hat{\beta}_i,\hat{\beta}_j) = [V(\hat{\beta}_i + \hat{\beta}_j) - V(\hat{\beta}_i) - V(\hat{\beta}_j)]/2.$$

For robust Wald tests of more than one restriction, a quadratic form needs to be constructed. For the null hypothesis

$$(2.31) \qquad H_0: R\beta = r,$$

where R is a QxK matrix, $Q \leq K$, rank(R) = Q, and r is a Qx1 vector, the Wald statistic is given by

$$(2.32) \qquad W = \sqrt{T}(R\hat{\beta} - r)'[R\hat{V}R']^{-1}\sqrt{T}(R\hat{\beta} - r)$$
$$= (R\hat{\beta} - r)'[R(\hat{V}/T)R']^{-1}(R\hat{\beta} - r),$$

where $\hat{V}$ is generally given by $(X'X/T)^{-1}\hat{B}(X'X/T)^{-1}$, and $\hat{B}$ is chosen to be heteroskedasticity or H/SC-consistent, as needed. Note that the correct formula for the Wald statistic is obtained by naively treating $\hat{\beta}$ is if it were distributed exactly as $N(\beta,\hat{V}/T)$. Under H_0,

$$(2.33) \qquad W \overset{a}{\sim} \chi_Q^2.$$

Under homoskedasticity and no serial correlation $\hat{V}$ can be taken to be $\hat{\sigma}^2(X'X/T)^{-1}$, where $\hat{\sigma}^2$ is given by (2.12). Plugging this choice of $\hat{V}$ into (2.30) and rearranging yields

$$(2.34) \qquad W = (R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/\hat{\sigma}^2$$
$$= QF,$$

where F is the standard F-statistic for testing (2.1). Under (2.10) and (2.11), F can be used as distributed approximately as $\mathcal{F}_{Q,N-K}$ (this is because $\mathcal{F}_{Q,N-K} \overset{d}{\to} \chi_Q^2/Q$ as $N \to \infty$). In general, the usual F-statistic does not have a known limiting distribution in the presence of conditional heteroskedasticity or serial correlation. In these cases the robust forms of $\hat{V}$ should be used.

3. Regression-Based Specification Tests

Regression-based diagnostics, which are frequently interpreted as Lagrange multiplier (LM) tests, are quite popular in time series econometrics. Pagan and Hall (1983) refer to such procedures as "residual analysis" because the statistics are motivated by examining the residuals from (in this case) an OLS regression. For example, consider testing the hypothesis H_0: $\gamma = 0$ in the model

$$E(y_t|x_t,\psi_t) = x_t\beta + \psi_t\gamma,$$

where ψ_t is a 1xQ subvector of $(1,z_t,y_{t-1},z_{t-1},\ldots)$ with lag lengths not depending on t. As in section 2, x_t generally denotes a 1xK subvector from $(1,z_t,y_{t-1},z_{t-1},\ldots)$ with lag lengths not depending on t. Under H_0,

(3.1) $$E(\psi_t' u_t) = 0,$$

where $u_t \equiv y_t - x_t\beta$ are the true errors under the null. The obvious way to operationalize (3.1) is to obtain the OLS residuals $\hat{u}_t$ from the regression

$$y_t \quad \text{on} \quad x_t, \quad t=1,\ldots,T$$

and check to see whether the sample covariance between ψ_t and $\hat{u}_t$,

(3.2) $$T^{-1}\sum_{t=1}^{T}\psi_t'\hat{u}_t,$$

is significantly different from zero. This is, in effect, what the Wald statistic for testing H_0: $\gamma = 0$ does, but it is possible to derive a statistic directly from (3.2). What is needed is the asymptotic variance of

(3.3) $$T^{-1/2}\sum_{t=1}^{T}\psi_t'\hat{u}_t = T^{-1/2}\sum_{t=1}^{T}\psi_t' u_t - T^{-1}\sum_{t=1}^{T}\psi_t' x_t\sqrt{T}(\hat{\beta} - \beta),$$

where $\hat{\beta}$ is the OLS estimator of β. Depending on the assumptions imposed under H_0, there are various ways that (3.3) can be used to derive a test statistic. Before proceeding further, it is useful to allow for a broader class of specification tests, as this is obtained without much additional work. Assume generally that the null hypothesis can be expressed as

(3.4) $$H_0\text{: } E(y_t|x_t,\psi_t) = x_t\beta,$$

where ψ_t is a 1xM vector of elements from $(z_t, y_{t-1}, z_{t-1}, \ldots)$. To allow for tests of neglected nonlinearity and endogeneity, let $\lambda(\psi_t, \eta)$ be a 1xQ vector of "misspecification indicators", which is allowed to depend on a vector of unknown nuisance parameters η. The choice of $\lambda(\psi_t, \eta)$ depends on the alternatives against which one would like to have power, and can contain linear and nonlinear functions of ψ_t. Several choices for λ will be discussed in the examples. The parameters η are called nuisance parameters because they need not have an interpretation as "true" parameters under H_0, although η is frequently equal to β.

The null hypothesis can be stated equivalently as

$$E(u_t | x_t, \psi_t) = 0 \tag{3.5}$$

where

$$u_t \equiv y_t - x_t\beta.$$

If $\hat{\eta}$ is an estimator such that $\sqrt{T}(\hat{\eta} - \eta) = O_p(1)$ then a test of (3.4) is based on

$$T^{-1} \sum_{t=1}^{T} \hat{\lambda}_t' \hat{u}_t,$$

where the $\hat{u}_t$ are the OLS residuals and $\hat{\lambda}_t \equiv \lambda(\psi_t, \hat{\eta})$ are the estimated misspecification indicators. Under (3.5) and standard regularity conditions, a simple mean value expansion shows that

$$T^{-1/2} \sum_{t=1}^{T} \hat{\lambda}_t' \hat{u}_t = T^{-1/2} \sum_{t=1}^{T} \lambda_t' \hat{u}_t + o_p(1), \tag{3.6}$$

where $\lambda_t \equiv \lambda(\psi_t, \eta)$ (e.g. Wooldridge (1990a)). This shows that, under H_0, the asymptotic distribution of $\hat{\eta}$ does not affect the asymptotic distribution of

$$T^{-1/2} \sum_{t=1}^{T} \hat{\lambda}_t' \hat{u}_t, \tag{3.7}$$

as long as $\hat{\eta}$ is $\sqrt{T}$-consistent for η. A more convenient form of (3.7) is

(3.8) $$T^{-1/2}\sum_{t=1}^{T}(\hat{\lambda}_t - x_t\hat{C})'\hat{u}_t \equiv \sum_{t=1}^{T}\hat{r}_t'\hat{u}_t,$$

where

$$\hat{C} \equiv \left(\sum_{t=1}^{T}x_t'x_t\right)^{-1}\sum_{t=1}^{T}x_t'\hat{\lambda}_t$$

is the KxQ matrix of regression coefficients from the regression

$$\hat{\lambda}_t \quad \text{on} \quad x_t, \quad t=1,\ldots,T,$$

and $\hat{r}_t$, t=1,...,T, are the 1xQ residual vectors from this regression. Let $\hat{\xi}_t \equiv \hat{u}_t\hat{r}_t$ and $\xi_t \equiv u_t r_t \equiv u_t(\lambda_t - x_tC)$, where $C \equiv \text{plim } \hat{C} = [E(x_t'x_t)]^{-1}E(x_t'\lambda_t)$, and $u_t \equiv y_t - x_t\beta$. Note that ξ_t is simply the population analog of $\hat{\xi}_t$: the estimated quantities $\hat{\beta}$, $\hat{\eta}$, and $\hat{C}$ have been replaced by their plims. The process $\{\hat{\xi}_t: t=1,2,\ldots,T\}$ has the useful property that

(3.9) $$T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t' - T^{-1/2}\sum_{t=1}^{T}\xi_t' \overset{p}{\to} 0$$

under H_0. To see this, note that

$$T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t' = T^{-1/2}\sum_{t=1}^{T}(\hat{\lambda}_t - x_tC)'\hat{u}_t$$

$$= T^{-1/2}\sum_{t=1}^{T}(\lambda_t - x_tC)'\hat{u}_t + o_p(1) \quad \text{by (3.6)}$$

$$= T^{-1/2}\sum_{t=1}^{T}(\lambda_t - x_tC)'u_t - T^{-1}\sum_{t=1}^{T}(\lambda_t - x_tC)'x_t\sqrt{T}(\hat{\beta} - \beta) + o_p(1)$$

$$= T^{-1/2}\sum_{t=1}^{T}(\lambda_t - x_tC)'u_t + o_p(1)\cdot O_p(1) + o_p(1)$$

since $E[(\lambda_t - x_tC)'x_t)] = 0$ and $\sqrt{T}(\hat{\beta} - \beta) = O_p(1)$. This establishes (3.9), and shows that asymptotic distribution of $T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t$ under H_0 is obtained once the easier problem of finding the asymptotic distribution of $T^{-1/2}\sum_{t=1}^{T}\xi_t$ has been solved. Equation (3.8) also makes it clear that the test based on

(3.5) is really a test of

$$H_0: E[(\lambda_t - x_t C)' u_t] = 0.$$

Because the $\hat{u}_t$ are orthognal to x_t by construction, the test checks whether the part of λ_t which is uncorrelated with x_t is correlated with u_t.

A test of (3.5) can be constructed for most stationary vector processes $\{\xi_t: t=1,2,\ldots\}$. If

$$\Xi \equiv \lim_{T\to\infty} V\left(T^{-1/2}\sum_{t=1}^{T}\xi_t\right) = E(\xi_t'\xi_t) + \sum_{j=1}^{\infty}\{E(\xi_t'\xi_{t+j}) + E(\xi_{t+j}'\xi_t)\}$$

is nonsingular, and if the central limit holds for $\{\xi_t\}$, then

$$\left(T^{-1/2}\sum_{t=1}^{T}\xi_t\right)\Xi^{-1}\left(T^{-1/2}\sum_{t=1}^{T}\xi_t\right)' \xrightarrow{d} \chi_Q^2$$

under H_0. If $\hat{\Xi}$ is a consistent estimator of Ξ, e.g.

$$\hat{\Xi} \equiv T^{-1}\sum_{t=1}^{T}\hat{\xi}_t'\hat{\xi}_t + \sum_{j=1}^{G}\varphi(j,G)\cdot T^{-1}\sum_{t=j+1}^{T}[\hat{\xi}_t'\hat{\xi}_{t-j} + \hat{\xi}_{t-j}'\hat{\xi}_t] \tag{3.10}$$

with $G = o(T^{1/4})$ and $\varphi(j,G)$ given by (2.23), then a computable statistic is

$$\left(T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t\right)\hat{\Xi}^{-1}\left(T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t\right)'. \tag{3.11}$$

The statistic (3.11) has an asymptotic χ_Q^2 distribution under H_0, and allows $E(u_t^2|x_t,\psi_t)$ and $E(u_{t+j}u_t|x_{t+j},\psi_{t+j},x_t,\psi_t)$ to be of fairly arbitrary form. Thus, (3.11) is one possible approach to computing a test statistic that is robust to heteroskedasticity and serial correlation.

Frequently it is useful to have available a statistic that can be computed via OLS regressions. It turns out that such a statistic can be derived which is still heteroskedasticity and serial correlation robust. The idea is simple: if ξ_t were a VAR(G) process (which is necessarily stable because of the ergodicity assumption), i.e.

$$(3.12) \qquad \xi_t = \xi_{t-1}R_1 + \xi_{t-2}R_2 + \ldots + \xi_{t-G}R_G + \nu_t,$$

where $\{\nu_t\}$ is a sequence of 1xQ uncorrelated errors, then $E(\xi_t) = 0$ if and only if $E(\nu_t) = 0$. A test of H_0 can be based on

$$(3.13) \qquad \left[T^{-1/2}\sum_{t=1}^{T}\nu_t\right]\left[T^{-1}\sum_{t=1}^{T}\nu_t'\nu_t\right]^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\nu_t\right]',$$

which has an asymptotic χ_Q^2 distribution under H_0. To operationalize (3.13), ν_t can be estimated as the residuals $\hat{\nu}_t$ from the vector autoregression

$$\hat{\xi}_t \text{ on } \hat{\xi}_{t-1}, \ldots, \hat{\xi}_{t-G}.$$

To justify replacing ν_t with $\hat{\nu}_t$ in (3.13), note that

$$T^{-1/2}\sum_{t=1}^{T}\hat{\nu}_t - T^{-1/2}\sum_{t=1}^{T}\nu_t = \sum_{j=0}^{G} T^{-1/2}\sum_{t=1}^{T}(\xi_{t-j}R_j - \hat{\xi}_{t-j}\hat{R}_j)$$

where $R_0 \equiv \hat{R}_0 \equiv I_Q$ and $\hat{R}_j$ are the QxQ coefficient matrices from the VAR. For each $j = 0,\ldots,G$,

$$(3.14) \qquad T^{-1/2}\sum_{t=1}^{T}(\xi_{t-j}R_j - \hat{\xi}_{t-j}\hat{R}_j) = T^{-1/2}\sum_{t=1}^{T}(\xi_{t-j}R_j - \xi_{t-j}\hat{R}_j)$$

$$+ T^{-1/2}\sum_{t=1}^{T}(\xi_{t-j}\hat{R}_j - \hat{\xi}_{t-j}\hat{R}_j)$$

$$= T^{-1/2}\sum_{t=1}^{T}\xi_{t-j}(R_j - \hat{R}_j)$$

$$+ T^{-1/2}\sum_{t=1}^{T}(\xi_{t-j} - \hat{\xi}_{t-j})\hat{R}_j.$$

But under H_0: $E(\xi_t) = 0$,

$$T^{-1/2}\sum_{t=1}^{T}(\xi_{t-j} - \hat{\xi}_{t-j}) = o_p(1)$$

and, by the CLT,

$$T^{-1/2}\sum_{t=1}^{T}\xi_{t-j} = O_p(1).$$

Under standard regularity conditions, $\hat{R}_j \xrightarrow{p} R_j$, and so $\hat{R}_j = O_p(1)$ under H_0. Thus, both terms on the right hand side of (3.14) are $o_p(1)$ under H_0, and it follows that

$$T^{-1/2}\sum_{t=1}^{T}\hat{\nu}_t - T^{-1/2}\sum_{t=1}^{T}\nu_t = o_p(1)$$

under H_0. A valid test statistic is

$$\left(T^{-1/2}\sum_{t=1}^{T}\hat{\nu}_t\right)\left(T^{-1}\sum_{t=1}^{T}\hat{\nu}_t'\hat{\nu}_t\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T}\hat{\nu}_t\right)',$$

which has an asymptotic χ^2_Q distribution under H_0. This statistic is easily seen to be $TR_u^2 = T - SSR$ from the regression

$$1 \text{ on } \hat{\nu}_t, \quad t=1,\ldots,T.$$

To summarize, the heteroskedasticity/serial correlation-robust procedure is

PROCEDURE 3.1:

(i) Obtain $\hat{u}_t$ as the residuals from the OLS regression

$$y_t \text{ on } x_t;$$

Compute the 1xQ vector indicator $\hat{\lambda}_t \equiv \lambda(\psi_t,\hat{\eta})$.

(ii) Obtain $\hat{r}_t$ as the 1xQ vectors of residuals from the regression

$$\hat{\lambda}_t \text{ on } x_t.$$

(iii) Define $\hat{\xi}_t$ to be the 1xQ vector $\hat{\xi}_t \equiv \hat{u}_t \cdot \hat{r}_t$. Save the 1xQ residuals $\hat{\nu}_t$ from the VAR(G) regression

$$\hat{\xi}_t \text{ on } \hat{\xi}_{t-1}, \ldots, \hat{\xi}_{t-G}.$$

(iv) Use $TR_u^2 = T - SSR$ from the regression

$$1 \text{ on } \hat{\nu}_t$$

as asymptotically χ^2_Q under H_0. In practice, one uses as T the actual number of observations used in this final regression. ■

The only step which is not automatic in Procedure 3.1, besides choosing the misspecification indicator $\hat{\lambda}_t$ (which will be discussed shortly), is the choice of G in step (iii). This is conceptually the same problem as choosing G in computing the Newey-West or Gallant-White covariance matrix estimator, and differs depending on the problem. The choice of G might depend on the frequency of the data and can differ across misspecification indicators. The key is to choose G so that $\{\nu_t\}$ is approximately uncorrelated. But if G is chosen too large relative to T, the chi-square distribution may not be a good approximation to the distribution of the test statistic.

It must be emphasized that Procedure 3.1 is <u>not</u> the same as assuming that the errors from the original model $\{u_t\}$ follow an AR(G) process and then computing $\hat{\beta}$ and test statistics based on a Cochrane-Orcutt type procedure. Such a procedure imposes strict exogeneity of x_t and common factor restrictions on the dynamic regression $E(y_t|x_t,\phi_{t-1})$, which are not necessarily intended under H_0. The VAR(G) in step (iii) is used to obtain estimates of ν_t. As long as G is selected appropriately, $\hat{\nu}_t$ will be approximately uncorrelated and step (iv) produces a valid test statistic. If G is too small then $T^{-1}\sum_{t=1}^{T}\hat{\nu}'_t\hat{\nu}_t$ is an inconsistent estimate of the variance of $T^{-1/2}\sum_{t=1}^{T}\nu_t$, but the statistic still has a well-defined limiting distribution under H_0; in contrast, a Cochrane-Orcutt type procedure could inappropriately reject H_0 with probability going to one.

There are two cases where Procedure 3.1 can be simplified. The first is when the null hypothesis imposes homoskedasticity and no serial correlation, applied now to u_t and (x_t,ψ_t). In other words, (2.10) and (2.11) are replaced by $E(u_{t+j}u_t|x_{t+j},\psi_{t+j},x_t,\psi_t) = 0$, $j \geq 1$ and $E(u_t^2|x_t,\psi_t) = \sigma^2$, respectively. In this case, using the fact that λ_t is a function of ψ_t,

$$\begin{aligned} E(\xi_t'\xi_t) &= E[E(\xi_t'\xi_t|x_t,\psi_t)] \\ &= E[E\{u_t^2(\lambda_t - x_tC)'(\lambda_t - x_tC)|x_t,\psi_t\}] \\ &= E[E(u_t^2|x_t,\psi_t)(\lambda_t - x_tC)'(\lambda_t - x_tC)] \\ &= \sigma^2 E[(\lambda_t - x_tC)'(\lambda_t - x_tC)] \\ &\equiv \sigma^2 E(r_t'r_t). \end{aligned}$$

Also,

$$\begin{aligned} E(\xi_{t+j}'\xi_t) &= E[E(\xi_{t+j}'\xi_t|x_{t+j},\psi_{t+j},x_t,\psi_t)] \\ &= E[E\{u_{t+j}u_t r_{t+j}'r_t|x_{t+j},\psi_{t+j},x_t,\psi_t\}] \\ &= E[E\{u_{t+j}u_t|x_{t+j},\psi_{t+j},x_t,\psi_t\}r_{t+j}'r_t] \\ &= 0. \end{aligned}$$

Under homoskedasticity and no serial correlation, Ξ has the very simple form

$$\Xi = E(\xi_t'\xi_t) = \sigma^2 T^{-1}\sum_{t=1}^{T} E(r_t'r_t),$$

and is consistently estimated by

$$\hat{\sigma}^2 T^{-1}\sum_{t=1}^{T} \hat{r}_t'\hat{r}_t$$

where $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2$. If this expression is used for $\hat{\Xi}$ in (3.11), the resulting statistic has a limiting χ_Q^2 distribution under H_0 and the additional assumptions of homoskedasticity and no serial correlation. It turns out that step (ii) of Procedure 3.1 is no longer necessary; instead, the statistic is computed as TR_u^2 from the regression

(3.15) $$\hat{u}_t \text{ on } x_t,\ \hat{\lambda}_t;$$

if x_t contains a constant then $\hat{u}_t$ has zero sample average and then the usual r-squared can be used as R_u^2. This form of the LM statistic is well-known and has been discussed by many authors; for examples and many references, see Pagan and Hall (1983) and Engle (1984). The computational simplicity of regression (3.15) is somewhat offset by its nonrobustness to heteroskedasticity and/or serial correlation. Because it is by far the most popular form of the LM statistic used its properties and limitations should be understood.

The homoskedasticity assumption can never be guaranteed to hold under the null as it concerns $V(y_t|x_t,\psi_t)$, not $E(y_t|x_t,\psi_t)$. On the other hand, in section 2 it was shown that the absense of serial correlation is a consequence of $E(y_t|x_t)$ being dynamically correctly specified. If the null hypothesis is

$$H_0\colon E(y_t|x_t) = E(y_t|x_t,\phi_{t-1}) \tag{3.16}$$

then, for any subvector ψ_t of elements from (x_t,ϕ_{t-1}),

$$E(u_{t+j}u_t|x_{t+j},\psi_{t+j},x_t,\psi_t) = 0 \tag{3.17}$$

by the law of iterated expectations. Consequently, if the null hypothesis imposes (3.16) either explicitly or implicitly, then there is no need to make the tests robust to serial correlation. Two examples of this are testing for Granger causality and testing for serial correlation; both of these take the null model to be dynamically complete. Obtaining tests for dynamic misspecification that are robust to heteroskedasticity is accomplished by simplifying Procedure 3.1. Because $\{\xi_t\}$ is serially uncorrelated under (3.16), Ξ is consistently estimated by

$$T^{-1}\sum_{t=1}^{T}\hat{\xi}_t'\hat{\xi}_t, \tag{3.18}$$

whether or not $E(u_t^2|x_t,\psi_t)$ is constant. Using (3.18) as $\hat{\Xi}$ in (3.11) is the same as skipping step (iii) in Procedure 3.1 and going directly to step (iv) with $\hat{\xi}_t$ in place of $\hat{\nu}_t$. Thus, the heteroskedasticity-robust LM statistic is obtained by performing steps (i), (ii), and

(iii′) Use $TR_u^2 = T - SSR$ from the regression

$$1 \text{ on } \hat{\xi}_t$$

as asymptotically χ_Q^2 under (3.5) and (3.17). Again, T here corresponds to the actual number of observations used in this last regression. Procedure (i), (ii), and (iii′) is robust in the presence of heteroskedascity, and loses nothing asymptotically in the event that heteroskedasticity is not present (Wooldridge (1990a)).

EXAMPLE 3.1 (Omitted Variables in a Static Regression Model): Consider the model

$$(3.19) \qquad y_t = \alpha + z_t\delta + \psi_t\gamma + u_t, \quad E(u_t|z_t,\psi_t) = 0,$$

where the 1xQ vector ψ_t is, like z_t, a set of contemporaneous variables. Interest lies in testing H_0: $\gamma = 0$ or $E(y_t|z_t,\psi_t) = E(y_t|z_t)$. Nothing guarantees that u_t will be homoskedastic or serially uncorrelated under H_0; the testable implication of H_0 is $E(\psi_t' u_t) = 0$. If interest lies in testing exclusion of ψ_t in a serial correlation robust manner then Procedure 3.1 can be applied by setting $x_t \equiv (1,z_t)$ and $\lambda(\psi_t,\eta) \equiv \psi_t$. The residuals $\hat{u}_t$ are obtained under H_0 from the regression

$$y_t \text{ on } 1,\ z_t,$$

and the $\hat{r}_t$ are obtained from the regression ψ_t on 1, z_t. An H/SC Chow test is obtained by setting $\psi_t \equiv (d_t,d_t z_t)$, where d_t is a dummy variable equal to unity after the hypothesized break point. The same procedure works if x_t and ψ_t are replaced by more general regressors. ∎

EXAMPLE 3.2 (Testing Functional Form in a Static Regression Model): Suppose that H_0 is specified generally as

$$H_0: E(y_t|z_t) = \alpha + z_t\delta \equiv x_t\beta.$$

A test for nonlinearities can be obtained, for example, by choosing $\lambda(\psi_t,\eta) \equiv ((x_t\beta)^2,(x_t\beta)^3)$, as in Ramsey's (1969) RESET. Then $Q = 2$, $\hat{\eta} = \hat{\beta}$, and $\hat{\lambda}_t = (\hat{y}_t^2,\hat{y}_t^3)$, where $\hat{y}_t$ are the fitted values from the regression y_t on 1, z_t; the $\hat{u}_t$ are obtained from the same regression. Note that $\psi_t \equiv x_t \equiv (1,z_t)$, and nothing guarantees that u_t is homoskedastic or serially uncorrelated under H_0. Other functions of z_t can be used in $\hat{\lambda}_t$, such as the fitted values from a nonlinear regression. Also, the same procedure applies to the general model

$$E(y_t|x_t) = x_t\beta,$$

where x_t can contain lagged values of y_t and/or z_t. ■

EXAMPLE 3.3 (Testing for Additional Lags in a Distributed Lag Model): Consider the finite DL model

$$E(y_t|z_t,z_{t-1},\ldots) = \alpha + z_t\delta_0 + \ldots + z_{t-M}\delta_M,$$

where z_t is a scalar for simplicity. For $P < M$, suppose the hypothesis of interest is

$$H_0: \delta_{P+1} = 0, \ldots, \delta_M = 0.$$

The number of restrictions is $Q = M - P$, $x_t \equiv (1,z_t,z_{t-1},\ldots,z_{t-P})$, $\hat{\lambda}_t \equiv \psi_t \equiv (z_{t-P-1},\ldots,z_{t-M})$, and $\hat{u}_t$ is obtained from the restricted regression

$$y_t \text{ on } 1,\ z_t,\ \ldots,\ z_{t-P}.$$

Again, the heteroskedasticity and serial correlation robust test is appropriate here, as nothing ensures that the errors u_t are serially uncorrelated under H_0. ■

EXAMPLE 3.4 (Testing for Serial Correlation in a General Dynamic Model): Suppose that under H_0,

$$y_t = x_t\beta + u_t, \quad E(u_t|x_t,\phi_{t-1}) = 0.$$

As mentioned above, a heteroskedasticity-robust LM statistic for AR(Q) serial correlation is obtained via steps (i), (ii), and (iii′) with $\hat{\lambda}_t \equiv (\hat{u}_{t-1},\ldots,\hat{u}_{t-Q})$ and $\hat{u}_t$ obtained from regressing y_t on x_t (note carefully the subscripts on the lagged residuals comprising $\hat{\lambda}_t$). ■

EXAMPLE 3.5 (Testing for Granger Causality): The null hypothesis in this case is

$$E(y_t|y_{t-1},z_{t-1},y_{t-2},z_{t-2},\ldots) = E(y_t|y_{t-1},y_{t-2},\ldots),$$

and to operationalize this it is assumed further that

$$E(y_t|y_{t-1},y_{t-2},\ldots) = E(y_t|y_{t-1},\ldots,y_{t-P}) = \alpha + \delta_1 y_{t-1} + \ldots + \delta_P y_{t-P}.$$

The lag length P needs to be selected, usually by choosing a value and then testing for additional lags of y. To test for Granger Causality, let $\hat{u}_t$ be the residuals from the regression

$$y_t \quad \text{on} \quad 1,\ y_{t-1},\ \ldots,\ y_{t-P}.$$

Then set $\hat{\lambda}_t \equiv (z_{t-1},\ldots,z_{t-Q})$ for some $Q \geq 1$, and use this either in regression (3.15) or in the heteroskedasticity-robust procedure (i), (ii), and (iii′). Note that the choice of Q is entirely up to the researcher.

4. Some Considerations When Applying Specification Tests

The results of sections 2 and 3 provide simple procedures for performing inference in linear time series models with ergodic data. Because time series analyses differ in their goals, the manner in which the various tests in section 3 are applied can differ across applications. Choosing a sensible strategy first requires deciding which relationship(s) between y and z is of interest. If the goal is to estimate a model for $E(y_t|z_t,y_{t-1},z_{t-1},\ldots)$, $E(y_t|y_{t-1},z_{t-1},\ldots)$, or $E(y_t|y_{t-1},y_{t-2},\ldots)$, then all tests are either implicitly or explicitly tests of dynamic specification. Such is the case for tests for serial correlation or Granger Causality, as well as the tests for common factor restrictions discussed below. Computation of the specification tests is simplified in this case because they need not be made serial correlation robust. It makes sense to compute both the standard form of the tests (either the F-test or LM test (3.15)) and the heteroskedasticity-robust LM test developed in section 3. Dynamic forms of heteroskedasticity are often found in regressions with financial data series, so the heteroskedasticity-robust forms might be particularly useful in testing asset pricing models.

There are certain problems for which the static expectation $E(y_t|z_t)$, the distributed lag expectation $E(y_t|z_t,z_{t-1},\ldots)$, or some other dynamically incomplete expectation is of interest. In this case one must distinguish among several null hypotheses. Godfrey (1987) has recently recommended a sequential specification testing strategy which attempts to test hypotheses in a logically consistent manner. The strategy suggested here is related to Godfrey's approach but differs in certain respects, including the form of the specification tests used.

First consider the case where the null hypothesis specifies a static linear model relating y_t to z_t. The first hypothesis of interest is the linearity of the conditional expectation $E(y_t|z_t)$. More formally, the hypothesis is

$$(4.1) \qquad E(y_t|z_t) = \alpha + z_t\delta \quad \text{(linearity)}.$$

A test of (4.1) can only be based on indicators that are nonlinear functions of z_t, say $\lambda(z_t,\hat{\eta})$ (e.g. see Example 3.2); to be robust to heteroskedasticity and serial correlation (neither of which can be ruled out if the null is (4.1)), Procedure 3.1 should be used.

A second hypothesis that is frequently of interest is whether an additional set of contemporaneous variables can be excluded from the linear model. If ψ_t is a 1xQ vector of contemporaneous variables (in addition to z_t), then the null hypothesis is

$$(4.2) \qquad E(y_t|z_t,\psi_t) = \alpha + z_t\delta \quad \text{(exclusion restrictions)}.$$

Assuming that the alternative to (4.2) is the linear model

$$(4.3) \qquad E(y_t|z_t,\psi_t) = \alpha + z_t\delta + \psi_t\gamma,$$

(4.2) is tested using the H/SC-robust Procedure 3.1 with $x_t \equiv (1,z_t)$ and $\hat{\lambda}_t \equiv \psi_t$. Applying Procedure 3.1 to both hypotheses (4.1) and (4.2) ensures that only the relevant nulls are assumed under H_0; $E(y_t|z_t,\phi_{t-1})$ and $V(y_t|z_t,\psi_t)$ are unrestricted up to regularity conditions.

It is important to stress that the hypotheses (4.1) and (4.2) are very different in that they restrict different conditional expectations: (4.1) restricts $E(y_t|z_t)$ while (4.2) restricts $E(y_t|z_t,\psi_t)$ (and hence $E(y_t|z_t)$). It is quite possible that (4.1) holds but (4.2) does not (e.g. if (y_t,z_t,ψ_t) are jointly normally distributed with $\gamma \neq 0$ in (4.3)). Further, if (4.1) holds then the tests for linearity of $E(y_t|z_t)$ discussed in Example 3.2 have no power for detecting violations of (4.2) (i.e. the asymptotic power is equal to the asymptotic size). Generally, RESET has power against $E(y_t|z_t,\psi_t) = \alpha + z_t\delta + \psi_t\gamma$, $\gamma \neq 0$, only if $E(\psi_t|z_t)$ is nonlinear. The only way to really test for omitted variables is to use those variables (ψ_t) in an LM or F-test. Although these comments are clearly illustrated when models are stated in terms of conditional expectations, there has been some confusion on this point in the literature (e.g. Thursby (1985)). The confusion arises when writing the model in error form and not accounting for the change in the coefficient on z_t when the conditioning set is reduced from (z_t,ψ_t) to z_t.

Next, one might want to test whether the static conditional expectation is equal to the DL expectation, i.e.

(4.4) $$E(y_t|z_t,z_{t-1},\ldots) = E(y_t|z_t) \quad \text{(no distributed lag dynamics).}$$

Again assuming linearity under H_0, this test is covered by Example 3.3 with P = 0 and M = Q to be chosen by the researcher. Again, Procedure 3.1 should be used because (4.4) implies nothing about $E(y_t|z_t,\phi_{t-1})$ or $V(y_t|z_t,z_{t-1},\ldots)$. As with hypotheses (4.1) and (4.2) the null model is $\alpha + z_t\delta$. But (4.4) is a hypothesis about a different expectation.

In some cases it might be hypothesized that the static expectation is in fact equal to the dynamic expectation:

(4.5) $$E(y_t|z_t,\phi_{t-1}) = E(y_t|z_t) \quad \text{(correct dynamic specification).}$$

Hypothesis (4.5) is of interest sometimes simply for the reason that the presence of serial correlation invalidates the use of the usual OLS test statistics. The most popular methods of testing (4.5) are (i) by testing for serial correlation in the errors, as in Example 3.4, or (ii) by forming an alternative model

(4.6) $$y_t = \alpha + z_t\delta + \psi_t\gamma + u_t,$$

where ψ_t now contains lagged values of y_t and/or z_t, and testing for exclusion of ψ_t by an F or LM test. The heteroskedasticity-robust LM test is obtained with $\hat{\lambda}_t \equiv \psi_t$ in steps (i), (ii), and (iii').

Hypotheses (4.1), (4.2), (4.4), and (4.5) represent restrictions on four different conditional expectations, even though the null specifies the same model. If the LM-type Procedure 3.1 and its variants are used, then all tests are based on the residuals $\hat{u}_t$ from the static regression

$$y_t \quad \text{on} \quad 1,\ z_t, \quad t=1,\ldots,T.$$

The misspecification indicator $\hat{\lambda}_t$ determines against which alternatives the test is likely to have power.

The analysis for a null finite distributed lag model is analogous to that for the static model. The null model is of the form $\alpha + z_t\delta_0 + z_{t-1}\delta_1 + \ldots + z_{t-P}\delta_P$. The analogs of (4.1), (4.2), (4.4), and (4.5) are almost immediate. For example, the null of correct dynamic specification is expressed as

$$E(y_t|z_t,\phi_{t-1}) = E(y_t|z_t,z_{t-1},\ldots).$$

This can be tested by including lags of y_t in the DL model and computing the F-statistic or by using the heteroskedasticity-robust LM procedure with $\hat{\lambda}_t$ containing lagged y_t.

The above analysis stresses that it is a good idea to compute serial correlation-robust standard errors when testing hypotheses about expectations other than $E(y_t|x_t,\phi_{t-1})$, for any 1xK subvector x_t. Nevertheless, because of its simplicity and proven usefulness, a more popular alternative for static or DL models is to use Cochrane-Orcutt methods to estimate an AR(1) process for the errors. If the AR(1) assumption is correct then this leads not only to consistent estimates of the coefficients in the static or DL model, but also to more efficient estimates. The static case where z_t is 1xJ is given by equations (2.16) and (2.17).

There are two essentially distinct tests that one can perform on the static/AR(1) model. The first focuses on the dynamic regression and the common factor restrictions. Recall from section 2 that the common factor restrictions (2.21) impose J nonlinear restrictions on the parameters of the dynamic expectation $E(y_t|z_t,\phi_{t-1})$. Because (2.21) is necessary for (2.16) and (2.17) to hold, a rejection implies that the usual statistics based on quasi-differenced regressions are invalid. And, of course, if one is interested only in $E(y_t|z_t,\phi_{t-1})$, then the nonlinear constraints imposed on the parameters of this expectation should be justified.

One way to test these restrictions on the dynamic regression model is to estimate the unrestricted vector $(\alpha_0,\gamma_0',\rho_1,\gamma_1')'$ by the OLS regression

$$y_t \text{ on } 1,\ z_t,\ y_{t-1},\ z_{t-1},$$

and to form the Wald test for the J nonlinear restrictions, as in Sargan (1964). Because the null hypothesis is correct dynamic specification, there is no need to make the statistic robust to serial correlation; on the other hand, a heteroskedasticity-robust version may be warranted.

An LM test is easily computed if the model is estimated by NLS. It too tests the restrictions (2.21) in model (2.20), and has no direct bearing the consistency of C-O or NLS for δ in (2.16). A simple example was covered in section 2 where the common factor restrictions hold yet $\delta_0 \neq \delta$.

If the goal of testing the static/AR(1) model is to examine whether C-O estimates are consistent for δ, then a different strategy is needed. First, it is important to derive the weakest set of conditions under which C-O consistently estimates δ. Recall that u_t is defined by

$$u_t = y_t - E(y_t|z_t) = y_t - \alpha - z_t\delta \equiv x_t\beta. \tag{4.7}$$

Whether or not (2.17) is true, C-O consistently estimates ρ in the equation

$$E(u_t|u_{t-1}) = \rho u_{t-1}. \tag{4.8}$$

This is because the first stage estimator of β is the OLS estimator $\hat{\beta}$, which is consistent for β in (4.7); then the autoregression of $\hat{u}_t$ on $\hat{u}_{t-1}$ consistently estimates ρ given by (4.8). The important step is then obtaining an estimator of β from a regression on quasi-differenced data; in what follows let $(\tilde{\beta},\tilde{\rho})$ be the C-O estimators, which may or may not be iterated after the first quasi-differenced regression. Then

$$\tilde{\beta} = \left[T^{-1}\sum_{t=1}^{T}\tilde{x}_t'\tilde{x}_t\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\tilde{x}_t'\tilde{y}_t\right],$$

where $\tilde{x}_t \equiv x_t - \tilde{\rho}x_{t-1}$ and $\tilde{y}_t \equiv y_t - \tilde{\rho}y_{t-1}$. Straightforward algebra shows that

$$\tilde{y}_t = \tilde{x}_t\beta + u_t - \tilde{\rho}u_{t-1},$$

so that

$$\tilde{\beta} = \beta + \left[T^{-1}\sum_{t=1}^{T}\tilde{x}_t'\tilde{x}_t\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\tilde{x}_t'(u_t - \tilde{\rho}u_{t-1})\right]$$

$$= \beta - \left[T^{-1}\sum_{t=1}^{T}\tilde{x}_t'\tilde{x}_t\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}(x_{t-1}'u_t + x_t'u_{t-1})\right]\tilde{\rho} + o_p(1).$$

The last equality follows because $E(x_t'u_t) = E(x_{t-1}'u_{t-1}) = 0$. By stationarity and the weak law of large numbers, $\tilde{\beta} \xrightarrow{p} \beta$ if and only if $E[(x_{t-1} + x_{t+1})'u_t] = 0$; because $E(u_t) = 0$ under (4.7), the condition reduces to

$$E[(z_{t-1} + z_{t+1})'u_t] = 0. \tag{4.9}$$

Thus, along with (4.7), condition (4.9) is the one underlying consistency of C-O in a static regression model. (Equation (4.8) is taken to be definitional, because the conditional expectation can always be replaced by the linear projection operator.) Note that (4.9) is not the same as the exogeneity condition (2.19) unless

$$E(z'_{t-1}u_t) = 0 \tag{4.10}$$

is maintained. In many static regressions (4.10) is assumed to be true, otherwise one would probably estimated a DL model. If a static model is estimated under the belief that there are no DL dynamics, then it makes sense to separate the hypotheses (4.10) and (2.19). Violation of (4.10) affects ones interpretation of δ, while violation of (2.19) makes C-O generally inconsistent for δ.

Condition (4.9) (or (2.19)) formally illustrates the point made earlier: the common factor restrictions on the dynamic regression play no direct role in the consistency of C-O. This fact helps to explain why in certain applications OLS estimates and C-O estimates appear to be close, even though the common factor restrictions are rejected. Or, on the other hand, why the common factor restrictions can appear to be supported by the data, yet C-O produces substantially different estimates of β.

Condition (4.9) is also the condition underlying the differencing specification test proposed by Plosser, Schwert, and White (1983), which compares the OLS coefficients from the regression in levels to the coefficients from a regression with differenced data. PSW simply set $\rho = 1$ in computing $\tilde{\beta}$; in other words, $\tilde{\beta}$ is obtained from the regression

(4.11) $$\Delta y_t \text{ on } \Delta z_t.$$

As shown above, ρ can be set to any number (provided the data are stationary or trend-stationary) or estimated by C-O (in which case $|\tilde{\rho}| < 1$ with probability approaching 1); the condition sufficient for consistency is always (4.9). A regression using the quasi-differenced data can be used to obtain $\tilde{\beta}$.

A test of (4.9) can be derived using Hausman's approach but, because the C-O estimator cannot be guaranteed to be more efficient than OLS under (4.7) and (4.9), it is easier to construct a direct test based on the OLS residuals (usually one would compute $\tilde{\beta}$ anyway to see if it differs from $\hat{\beta}$ in an economically signifcant way). The test procedure is to simply estimate the model by OLS and use $\hat{\lambda}_t \equiv z_{t-1} + z_{t+1}$ as the misspecification indicator in Procedure 3.1. If the presence of distributed lag dynamics is a separate hypothesis, then (2.19) should be tested directly with $\hat{\lambda}_t \equiv z_{t+1}$. If the test rejects, the C-O estimates need not be computed because they are necessarily inconsistent; OLS should be used to consistently estimated β, and robust standard errors and test statistics can be used to perform inference.

The strength of this approach is that it tests only the assumption needed for C-O to be consistent, and provides insight into why the OLS and Cochrane-Orcutt estimates might be far apart. Unfortunately, a failure to reject only leads to further questions. While a failure to reject lends support for (2.19), one cannot be confident that (2.16) and (2.17) hold. Thus, although C-O might be consistent for δ, it does not necessarily have the other desirable properties usually associated with it (being more efficient than OLS, resulting in computationally simple test statistics).

In order to justify the use of the usual C-O statistics (possibly corrected for neglected heteroskedasticity), one should test the validity of the common factor restrictions as well as (2.19). Because (2.16) and (2.17) are difficult to relax in any useful way while maintaining the validity of statistics from C-O estimation, the common factor tests are derived under these assumptions.

To derive the LM test of common factor restrictions based on the C-O estimates, let let $\tilde{\alpha}$, $\tilde{\delta}$, and $\tilde{\rho}$ be the Cochrane-Orcutt estimators of α, δ, and ρ. The residuals from this estimation are

(4.12)
$$\begin{aligned}\tilde{e}_t &\equiv \tilde{u}_t - \tilde{\rho}\tilde{u}_{t-1}\\ &\equiv \tilde{y}_t - \tilde{x}_t\tilde{\beta}\end{aligned}$$

where $\tilde{u}_t \equiv y_t - x_t\tilde{\beta}$, $\tilde{y}_t \equiv y_t - \tilde{\rho}y_{t-1}$, and $\tilde{x}_t \equiv x_t - \tilde{\rho}x_{t-1}$. (The first observation can be treated in the usual way: $\tilde{y}_1 \equiv [1 - \tilde{\rho}^2]^{1/2}y_1$, $\tilde{x}_1 \equiv [1 - \tilde{\rho}^2]^{1/2}x_1 \equiv [1 - \tilde{\rho}^2]^{1/2}(1,z_1)$, $\tilde{u}_1 \equiv y_1 - x_1\tilde{\beta}$, $\tilde{e}_1 \equiv [1 - \tilde{\rho}^2]^{1/2}\tilde{u}_1$.) The gradient of the restricted regression function with respect to α, δ, and ρ evaluated at the estimates is

(4.13)
$$(x_t - \tilde{\rho}x_{t-1}, y_{t-1} - x_{t-1}\tilde{\beta}) \equiv (\tilde{x}_t, \tilde{u}_{t-1}).$$

The unrestricted gradient is simply $(1,z_t,y_{t-1},z_{t-1})$. The standard LM test is obtained as TR^2 from the regression

(4.14)
$$\tilde{e}_t \text{ on } 1,\ z_t,\ y_{t-1},\ z_{t-1},$$

or equivalently from

$$\tilde{e}_t \text{ on } 1,\ \tilde{z}_t,\ \tilde{u}_{t-1},\ z_{t-1}.$$

(When a trend is included in the original estimation, a trend is simply included in (4.14); just as the common factor restriction on the intercept cannot be tested, neither can those on polynomial trends.) Under the assumptions (2.16), (2.17), and conditional homoskedasticity, $TR^2 \overset{d}{\rightarrow} \chi_J^2$.

The heteroskedasticity-robust form is obtained by applying the results of Wooldridge (1990b) for nonlinear regression: first regress

(4.15) $$z_{t-1} \quad \text{on} \quad 1,\ \tilde{z}_t,\ \tilde{u}_{t-1}$$

and save the 1xJ residuals, say $\tilde{r}_t$. Then use $TR_u^2 = T - SSR$ from the regression

(4.16) $$1 \quad \text{on} \quad \tilde{e}_t\tilde{r}_t$$

as asymptotically χ_J^2 under H_0. This is completely analagous to steps (i), (ii), and (iii') in section 3. If time trends are included in the initial estimation, the same functions of time are included on the right hand side of (4.15).

These tests have immediate extensions for the finite distributed lag model of order Q with AR(1) errors:

(4.17) $$y_t = \alpha + z_t\delta_0 + \ldots + z_{t-Q}\delta_Q + u_t, \quad E(u_t|z_t,z_{t-1},\ldots) = 0$$

(4.18) $$u_t = \rho u_{t-1} + e_t, \quad E(e_t|z_t,\phi_{t-1}) = 0.$$

Under (4.17), the additional condition needed for C-O to consistently estimate $\delta_0,\ldots,\delta_Q$ is (2.19) (note that all lags of z are assumed to be uncorrelated with u, so that the analog of (4.9) reduces to (2.19)). Thus, the test is carried out as before: let $\hat{u}_t$ be the residuals from the OLS regression

$$y_t \quad \text{on} \quad 1,\ z_t,\ z_{t-1},\ \ldots,\ z_{t-Q},$$

let $x_t \equiv (1,z_t,\ldots,z_{t-Q})$, let $\hat{\lambda}_t \equiv z_{t+1}$, and use these in Procedure 3.1.

For the common factor test, let $\tilde{e}_t \equiv \tilde{u}_t - \tilde{\rho}\tilde{u}_{t-1}$, where $\tilde{u}_t \equiv y_t - \tilde{\alpha} - z_t\tilde{\delta}_0 - \ldots - z_{t-Q}\tilde{\delta}_Q$ and $\tilde{\alpha}$, $\tilde{\delta}_0$, ..., $\tilde{\delta}_Q$, $\tilde{\rho}$ are obtained from a Cochrane-Orcutt procedure. The usual LM test is simply TR^2 from

$$\tilde{e}_t \text{ on } 1, z_t, y_{t-1}, z_{t-1}, \ldots, z_{t-Q-1}, \tag{4.19}$$

which is asymptotically χ^2_J under H_0 and conditional homoskedasticity. The heteroskedasticity-robust test obtains $\tilde{r}_t$ as the 1xJ residuals from the regression

$$z_{t-Q-1} \text{ on } 1, \tilde{z}_t, \ldots, \tilde{z}_{t-Q}, \tilde{u}_{t-1}, \tag{4.20}$$

where $\tilde{z}_{t-j} \equiv z_{t-j} - \tilde{\rho}z_{t-j-1}$, j=0,...,Q, and uses them as in (4.15). Again, any time trend used in obtaining the C-O estimates should be used on the right hand side of (4.20).

One caveat about these tests. The test of (2.19) recommended here is not the best possible if (2.16) and (2.17) are maintained under H_0. In this case, it would be better to base a test on the C-O residuals rather than on the OLS residuals (or, construct a Hausman test which directly compares $\tilde{\delta}$ and $\hat{\delta}$). But the tests using $\hat{\lambda}_t \equiv z_{t+1}$ in Procedure 3.1 are robust in that they take only (2.19) to be the null. This test can be used to indicate whether C-O is leading one astray in terms of parameter estimates. The test for common factor restrictions, which have been derived under (2.16) and (2.17), can then be used to check the additional assumptions required for the validity of the statistics based on quasi-differenced data.

5. Results for Two Stage Least Squares

Again consider the linear model

$$y_t = x_t\beta + u_t, \quad t=1,2,\ldots, \tag{5.1}$$

where x_t is 1xK and y_t and u_t are scalars. However, suppose now that the parameters β do not index the conditional expectation $E(y_t|x_t)$ or, more traditionally, that some elements of x_t are correlated with u_t. This can be the case for a variety of reasons: (5.1) might be an equation in a simultaneous equations model where x_t contains jointly determined variables; x_t might contain proxies of the true variables of interest; or (5.1) might omit variables that one would like to control for. In such cases there is a conditional expectation that one would like to estimate but simultaneity, sample selection, errors in variables, or unobserved variables makes it impossible to do so by an OLS regression of y_t on x_t.

Instead, let w_t be a set of 1xL instrumental variables chosen from (z_t, ϕ_{t-1}) with $L \geq K$; the restriction on w_t is

$$E(u_t|w_t) = 0; \tag{5.2}$$

for some of the subsequent analysis (5.2) can be replaced by the zero correlation assumption $E(w_t'u_t) = 0$, but for clarity (5.2) is assumed to be in force throughout. The vector w_t excludes any elements of z_t that are simultaneously determined with y_t, but w_t would contain the elements of x_t that satisfy (5.2). Recall that the 2SLS estimator of β is

$$\begin{aligned} \hat{\beta} &\equiv (\hat{X}'\hat{X})^{-1}\hat{X}'Y = (\hat{X}'X)^{-1}\hat{X}'Y \\ &= \beta + (\hat{X}'\hat{X})^{-1}\hat{X}'U, \end{aligned} \tag{5.3}$$

where $\hat{X}$ is the TxK matrix with t^{th} row $\hat{x}_t$, and $\hat{x}_t \equiv w_t\hat{\Pi}$ is the 1xK vector of fitted values from the regression

$$x_t \text{ on } w_t, \quad t=1,\ldots,T.$$

Analogous to the OLS estimator, (5.2) is the crucial condition for the 2SLS estimator to be consistent for β. The errors u_t can contain fairly arbitrary forms of heteroskedasticity and serial correlation, and x_t and w_t can contain lagged dependent variables. Under standard regularity conditions the 2SLS estimator is asymptotically normally distributed.

By stationarity, the coefficients from the linear projection of x_t on w_t are time invariant: $x_t^* \equiv w_t\Pi$, where $\Pi \equiv [E(w_t'w_t)]^{-1}E(w_t'x_t)$. An important fact about the 2SLS estimator is that

$$\text{(5.4)} \qquad \sqrt{T}(\hat{\beta} - \beta) = (X^{*\prime}X^*/T)^{-1}T^{-1/2}X^{*\prime}U + o_p(1)$$
$$= A^{-1}T^{-1/2}X^{*\prime}U + o_p(1)$$

where $A \equiv E(x_t^{*\prime}x_t^*)$. Equation (5.4) shows that the fact the fitted values $\hat{x}_t$ have been estimated does <u>not</u> affect the limiting distribution of the 2SLS estimator; the same limiting distribution is obtained if $\hat{x}_t$ is replaced by x_t^*, i.e. if Π replaces $\hat{\Pi}$. This makes it easy to obtain consistent standard errors in a variety of circumstances. In the general case, the asymptotic covariance matrix of $\sqrt{T}(\hat{\beta} - \beta)$ is given by $V \equiv A^{-1}BA^{-1}$ where now

$$\text{(5.5)} \qquad B \equiv E(s_t^{*\prime}s_t^*) + \sum_{j=1}^{\infty}[E(s_{t+j}^{*\prime}s_t^*) + E(s_t^{*\prime}s_{t+j}^*)],$$

and $s_t^* \equiv x_t^*u_t$.

In the context of 2SLS, the assumption of no serial correlation is most easily stated as

$$\text{(5.6)} \qquad E(u_{t+j}u_t|w_{t+j},w_t) = 0, \; j \geq 1.$$

In (5.6), w_t can be replaced by x_t^* with the subsequent results going through. Technically, this allows for certain forms of serial correlation ruled out by (5.6), but the additional generality is quite modest.

The appropriate homoskedasticity assumption is

$$\text{(5.7)} \qquad E(u_t^2|w_t) = \sigma^2, \quad t=1,2,\ldots,$$

which imposes homoskedasticity of u_t conditional on the instruments w_t. Again, x_t^* can replace w_t in stating (5.7).

Under (5.6) and (5.7), the usual asymptotic covariance matrix estimator is consistent for avar $\sqrt{T}(\hat{\beta} - \beta)$. Let $\hat{u}_t \equiv y_t - x_t\hat{\beta}$ denote the 2SLS residuals. Then a consistent estimator of σ^2 is given by

$$\hat{\sigma}^2 \equiv (T-K)^{-1} \sum_{t=1}^{T} \hat{u}_t^2;$$

the degrees of freedom adjustment does not make a difference asymptotically, and is used by most regression packages. The asymptotic standard error of $\hat{\beta}_j$ is the square root of the jth diagonal element of

$$\hat{\sigma}^2(\hat{X}'\hat{X})^{-1}.$$

This is what is printed out by all regression packages.

In the present context, dynamic completeness is defined by

$$E(u_t|w_t,\phi_{t-1}) = 0, \tag{5.8}$$

where ϕ_{t-1} contains all past values of w, x, and y. As with the case of OLS, (5.8) can be shown to imply (5.6) by a standard application of the law of iterated expectations. Setting

$$\hat{B} \equiv \hat{X}'\hat{\Sigma}\hat{X}/(T-K) = (T-K)^{-1} \sum_{t=1}^{T} \hat{u}_t^2\hat{x}_t'\hat{x}_t,$$

where $\hat{\Sigma} \equiv \mathrm{diag}(\hat{u}_1^2,\ldots,\hat{u}_T^2)$, a heteroskedasticity-robust covariance matrix estimator is

$$\hat{V} \equiv (\hat{X}'\hat{X}/T)^{-1}[\hat{X}'\hat{\Sigma}\hat{X}/(T-K)](\hat{X}'\hat{X}/T)^{-1}, \tag{5.9}$$

and the asymptotic standard error of $\hat{\beta}_j$ is the square root of the jth diagonal element of $\hat{V}$.

For the general case that $\{w_t'u_t\}$ might be serially correlated and $E(u_t^2|w_t)$ nonconstant, let

$$\hat{\Omega}_j \equiv T^{-1} \sum_{t=j+1}^{T} \hat{s}_t' \hat{s}_{t-j}, \quad \hat{s}_t \equiv \hat{x}_t \hat{u}_t,$$

and compute an estimator of B by (2.22). The only difference between this case and the OLS case is that x_t has been replaced by $\hat{x}_t$ everywhere except (as usual in 2SLS contexts) in the computation of $\hat{u}_t$. The asymptotic variance estimator of $\hat{\beta}$ is still given by

$$\hat{V}/T \equiv (\hat{X}'\hat{X})^{-1}(T\hat{B})(\hat{X}'\hat{X})^{-1}.$$

Procedure 3.1 has an immediate generalization for computing a heteroskedasticity-serial correlation robust standard error for $\hat{\beta}_j$:

PROCEDURE 5.1:

(i) Estimate $\hat{\beta}$ by 2SLS using instruments w_t. This yields "se($\hat{\beta}_j$)", $\hat{\sigma}$, and the 2SLS residuals $\{\hat{u}_t: t=1,\ldots,T\}$. Obtain the fitted values $\hat{x}_t$ from the first step regression

$$x_t \text{ on } w_t.$$

(ii) Compute the residuals $\{\hat{r}_{tj}: t=1,\ldots,T\}$ from the regression

(5.10) $$\hat{x}_{tj} \text{ on } \hat{x}_{t1},\ldots,\hat{x}_{t,j-1},\hat{x}_{t,j+1},\ldots,\hat{x}_{tK}, \quad t=1,\ldots,T$$

(iii) Set $\hat{\xi}_t \equiv \hat{r}_{tj}\hat{u}_t$ and run the regression

(5.11) $$\hat{\xi}_t \text{ on } \hat{\xi}_{t-1}, \ldots, \hat{\xi}_{t-G},$$

where G is, say, the integer part of $T^{1/4}$. Compute the spectral density estimator

$$\hat{c}_j \equiv [T/(T-K)]\hat{\tau}_G^2/(1 - \hat{a}_1 - \hat{a}_2 - \ldots - \hat{a}_G)^2,$$

where $\hat{a}_j$, j=1,...,G, are the OLS coefficients from the autoregression (5.11) and $\hat{\tau}_G^2$ is the square of the standard error of the regression.

(iv) Compute $se(\hat{\beta}_j)$ from

(5.12) $$se(\hat{\beta}_j) = ["se(\hat{\beta}_j)"/\hat{\sigma}]^2(T\hat{c}_j)^{1/2}.$$ ∎

The standard error from Procedure 5.1 is both heteroskedasticity and (as $T \rightarrow \infty$) serial correlation robust. Showing that this produces a consistent standard error follows along the lines of Wooldridge (1990b).

Regression-based specification tests require only a slight modification from the OLS case. As in Section 3, let ψ_t denote a set of "exogenous" and/or predetermined variables from (z_t, ϕ_{t-1}). Elements of x_t that are correlated with u_t are excluded by definition, but ψ_t can contain elements from w_t and other variables from ϕ_{t-1} that should be uncorrelated with u_t under H_0. The null hypothesis is taken to be

(5.13) $$H_0: E(u_t|w_t, \psi_t) = 0.$$

Let $\lambda(\psi_t, \eta)$ be a 1xQ vector of misspecification indicators with nuisance parameter estimator $\hat{\eta}$. The test of (5.18) is based on

$$T^{-1}\sum_{t=1}^{T}\hat{\lambda}_t'\hat{u}_t,$$

where $\hat{\lambda}_t \equiv \lambda(\psi_t, \hat{\eta})$ and $\hat{\eta}$ is a nuisance parameter estimator. The general H/SC-robust procedure is an immediate extension from Procedure 3.1.

PROCEDURE 5.2:

(i) Obtain $\hat{u}_t$ as the residuals from the 2SLS regression

$$y_t \quad \text{on} \quad x_t \quad \text{using instruments } w_t.$$

Compute the 1xK fitted values $\hat{x}_t$ from the first stage regression of x_t on w_t, or from x_t on w_t, $\hat{\lambda}_t$.

(ii) Obtain $\hat{r}_t$ as the 1xQ vectors of residuals from the regression

$$\hat{\lambda}_t \quad \text{on} \quad \hat{x}_t.$$

(iii) Define $\hat{\xi}_t$ to be the 1xQ vector $\hat{\xi}_t \equiv \hat{u}_t\hat{r}_t$. Save the 1xQ residuals $\hat{\nu}_t$ from the VAR(G) regression

$$\hat{\xi}_t \quad \text{on} \quad \hat{\xi}_{t-1}, \ \ldots, \ \hat{\xi}_{t-G}.$$

(iv) Use $TR_u^2 = T - SSR$ from the regression

$$1 \quad \text{on} \quad \hat{\nu}_t;$$

T can be the actual number of observations used in this final regression. ■

The choice of whether to compute $\hat{x}_t$ from the regression of x_t on w_t or x_t on w_t <u>and</u> $\hat{\lambda}_t$ depends on what is assumed about $E(x_t|w_t,\lambda_t)$ under H_0. To see the issue, note that

$$(5.14) \qquad T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t = T^{-1/2}\sum_{t=1}^{T}\hat{u}_t\hat{r}_t = T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\hat{\beta})(\hat{\lambda}_t - \hat{x}_t\hat{C})$$

where

$$\hat{C} \equiv (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{\Lambda}.$$

Underlying Procedure 5.2 is the assumption

$$(5.15) \qquad T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t = T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\beta)(\lambda_t - x_t^*C) + o_p(1),$$

where x_t^* are the population analogs of $\hat{x}_t$ and $C \equiv [E(x_t^{*\prime}x_t^*)]^{-1}E(x_t^{*\prime}\lambda_t)$; in other words, each estimator implicit in $\hat{\xi}_t$ can be replaced by its plim without altering the asymptotic distribution of its standardized partial sum. This was shown to always be the case for the OLS tests in Section 3. As in section 3, the fact that $\hat{\lambda}_t$ is estimated does not affect the limiting distribution of $T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t$ under H_0. Also, by the first order condition for the 2SLS estimator,

$$T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t = T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\hat{\beta})(\hat{\lambda}_t - \hat{x}_tC),$$

so

$$T^{-1/2}\sum_{t=1}^{T}\hat{\xi}_t = T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\hat{\beta})(\lambda_t - \hat{x}_tC) + o_p(1).$$

But

$$T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\hat{\beta})(\lambda_t - \hat{x}_tC) = T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\hat{\beta})(\lambda_t - x_t^*C)$$

$$+ T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\hat{\beta})w_t(\hat{\Pi} - \Pi)$$

$$= T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\hat{\beta})(\lambda_t - x_t^*C) + O_p(1)\cdot o_p(1)$$

under H_0. Consequently, (5.15) holds if it can be shown that

$$T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\hat{\beta})(\lambda_t - x_t^*C) = T^{-1/2}\sum_{t=1}^{T}(y_t - x_t\beta)(\lambda_t - x_t^*C) + o_p(1)$$

under H_0. Because $\sqrt{T}(\hat{\beta} - \beta) = O_p(1)$ and $E[(\lambda_t - x_t^*C)'u_t] = 0$ under H_0, it suffices to show that

(5.16) $$T^{-1}\sum_{t=1}^{T}x_t'(\lambda_t - x_t^*C) \overset{p}{\to} 0;$$

by the WLLN, (5.16) holds provided

(5.17) $$E[x_t'(\lambda_t - x_t^*C) = 0.$$

Suppose now that

(5.18) $$x_t^* = E(x_t|w_t,\lambda_t).$$

Then the law of iterated expectations implies that

(5.19) $$\begin{aligned} E[x_t'(\lambda_t - x_t^*C)] &= E[E(x_t|w_t,\lambda_t)'(\lambda_t - x_t^*C)] \\ &= E[x_t^{*\prime}(\lambda_t - x_t^*C)] \\ &= 0 \end{aligned}$$

by definition of C, verifying (5.17). Thus, the validity of Procedure 5.2 generally requires x_t^* to be the predicted values from the population regression of x_t on w_t and λ_t. However, in many cases, $E(x_t|w_t,\lambda_t) = E(x_t|x_t) = w_t\Pi$ under H_0 (otherwise λ_t would already be in the instrument list), in which case $\hat{x}_t$ can be obtained from the regression of x_t on w_t. If there is any doubt whether λ_t has additional predictive power for x_t under H_0, then the regression of x_t on w_t, $\hat{\lambda}_t$ can be used to obtain $\hat{x}_t$.

If the null hypothesis imposes (5.8) then then $E(u_{t+j}u_t|w_{t+j},\psi_{t+j},w_t,\psi_t) = 0$ for any $\psi_t \subset (w_t,\phi_{t-1})$, and the test need not be made robust to serial correlation. A heteroskedasticity-robust test is obtained by replacing (iii) and (iv) by

(iii′) Regress

(5.20) $$1 \quad \text{on} \quad \hat{\xi}_t, \; t=1,\ldots,T$$

and use T - SSR as asymptotically χ_Q^2.

A test which imposes $E(u_t^2|w_t,\psi_t) = \sigma^2$ and $E(u_{t+j}u_t|w_{t+j},\psi_{t+j},w_t,\psi_t) = 0$ is obtained as TR_u^2 from the regression

(5.21) $$\hat{u}_t \quad \text{on} \quad \hat{x}_t, \; \hat{\lambda}_t.$$

Both (5.21) and (5.20) require (5.18).

EXAMPLE 5.1 (Testing For Omitted Variables): Consider the model

(5.22) $$y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + u_t,$$

where the null hypothesis is

$$H_0: \beta_2 = 0.$$

Both x_{t1} and x_{t2} can contain elements correlated with u_t. In general, the list of valid instruments can change under the null and alternative models.

For example, if x_{t2} contains lagged endogenous variables then, in many cases, these lags would not be used as instruments if H_0 were true. Let w_t be a 1xL set of valid instruments under H_0. Assume that $L \geq K \equiv K_1 + K_2$. Let w_{t1} be a $1xL_1$ subvector of w_t such that $E(x_{t1}|w_t) = w_{t1}\Pi_1$; assume that $L_1 \geq K_1$. Let $\hat{\beta}_1$ and $\hat{u}_t$ denote the 2SLS statistics obtained using instruments w_{t1} under the restriction H_0: $\beta_2 = 0$, and let $\hat{x}_{t1}$ denote the fitted values from the first step regression x_{t1} on w_{t1}. (If x_{t1} is exogenous then $w_{t1} = x_{t1}$ and $\hat{x}_{t1} = x_{t1}$). Let $\hat{x}_{t2}$ be the fitted values from the regression x_{t2} on w_t. Then $\hat{u}_t$ and $\hat{\lambda}_t \equiv \hat{x}_{t2}$ can be used in Procedure 5.2. What is really being tested is whether $\text{plim}\ T^{-1}\sum_{t=1}^{T} \hat{x}'_{t2}\hat{u}_t = 0$. If the test is intended to detect dynamic misspecification then (5.20) or (5.21) can be used according to whether or not homoskedasticity is maintained. ■

EXAMPLE 5.2 (Testing for Serial Correlation): Let $\hat{u}_t$ and $\hat{x}_t$ be obtained from 2SLS estimation of

$$y_t \quad \text{on} \quad x_t \quad \text{using IV's } w_t.$$

A test for AR(Q) serial correlation is obtained by using $\hat{\lambda}_t \equiv (\hat{u}_{t-1}, \ldots, \hat{u}_{t-Q})$ in (i), (ii), and (iii′) or in regression (5.21) (not robust to heteroskedasticity). If $u_{t-1}, \ldots u_{t-Q}$ add explanatory power to x_t under H_0 then $\hat{x}_t$ should be obtained from x_t on w_t, $\hat{u}_{t-1}$, ..., $\hat{u}_{t-Q}$. ■

EXAMPLE 5.3 (Testing for Endogeneity): Let the model be partioned as in (5.22), where x_{t1} is taken to be exogenous. The issue is whether x_{t2} is endogenous:

$$H_0\colon E(u_t|x_{t2}) = 0.$$

The model is estimated by OLS under H_0, so let $\hat{u}_t$ denote the OLS residuals from the regression y_t on x_{t1}, x_{t2}. If w_t denotes a set of instruments that includes x_{t1} but not x_{t2} then a Hausman test which compares the OLS and 2SLS estimators can be shown to be based on the sample covariance

$$T^{-1} \sum_{t=1}^{T} \hat{x}'_{t2}\hat{u}_t,$$

where the $\hat{x}_{t2}$ are the fitted values from the first stage regression x_{t2} on w_t. Thus, take $\hat{x}_t \equiv x_t$ and $\hat{\lambda}_t \equiv \hat{x}_{t2}$ in Procedure 5.2; the degrees of freedom of the test is K_2, the dimension of x_{t2}. A test which assumes homoskedasticity and no serial correlation is based on TR_u^2 from the regression

$$\hat{u}_t \quad \text{on} \quad x_{t1}, x_{t2}, \hat{x}_{t2};$$

see Hausman (1983). Steps (i)-(iv) or (i)-(iii′) can be used to obtain robust versions. Note that, because the null model is estimated by OLS, this test also falls under Procedure 3.1. ■

EXAMPLE 5.4 (Testing Overidentifying Restrictions): Let the model be

$$y_t = x_t\beta + u_t,$$

where x_t is 1xK. Let w_t be a 1xL vector of instruments, where $L > K$. If $\hat{u}_t$ denotes the 2SLS residuals $y_t - x_t\hat{\beta}$, a test of overidentifying restrictions, which assumes homoskedasticity and no serial correlation under the null, is obtained as TR_u^2 from the regression $\hat{u}_t$ on w_t; TR_u^2 is asymptotically χ_Q^2, where $Q \equiv L - K$. Procedure 5.2 is applied by taking $\hat{x}_t = w_t\hat{\Pi}$ and $\hat{\lambda}_t$ any of Q elements from w_t that are not also elements of x_t. This produces an H/CS-robust test of the overidentifying restrictions. Steps (i), (ii), and (iii′) produce the heteroskedasticity-robust form. ■

6. Concluding Remarks

The procedures suggested in this paper offer relatively simple methods for carrying out inference in linear time series models that are robust to fairly arbitrary forms of serial correlation and heteroskedasticity. The standard errors and test statistics discussed in sections 2 - 5 are alternatives to more popular methods which model serial correlation in the errors and impose certain exogeneity assumptions on the regressors. The H/SC-robust forms of the test statistics require only very weak assumptions on the errors.

The observation that the very weak requirement $E(u_t|x_t) = 0$ (OLS) or $E(u_t|w_t) = 0$ (2SLS) suffices for consistency (along with regularity conditions) raises an interesting question which has not received much attention lately. Namely, what exactly should be required of the errors in time series models? If the errors should only be required to be uncorrelated with the regressors (OLS) or instruments (2SLS), then the methods of this paper have significant robustness advantages over more traditional serial correlation modelling approaches. If "correct specification" requires that the errors be serially uncorrelated (unforecastable), then many static and distributed lag models are necessarily misspecified. By this criterion most time series regressions would need to contain lags of dependent as well as lags of conditioning variables.

As mentioned in the introduction, many approaches to economic modelling do not allow one to address the question about what should be required of the errors. Most of the conditions imposed on the errors have arisen out of statistical considerations. In the context of the linear model, the no serial correlation assumption (2.10) (along with the homoskedasticity assumption (2.11)) validates the usual OLS test statistics, at least asymptotically. The static model with AR(1) errors, given by (2.16) and (2.17), originated primarily to obtain standard errors and test statistics with the usual properties; it also produces an estimator which is asymptotically more efficient than OLS.

It was some time later that econometricians realized that (2.16) and (2.17) impose common factor restrictions on the dynamic regression. This paper has further emphasized the additional exogeneity restriction (2.19). It seems useful to seek conditions on the errors in static and distributed lag models that have economic content, rather than being motivated by statistical considerations. One candidate approach is rational expectations, which imposes unforecastability given a certain information set. Unfortunately, many static relationships are estimated without appealing at all to rational expectations. A broader set of criteria is needed. One possible requirement, that seems to not have appeared in the literature, is that u be Granger causally prior to z, i.e.

$$E(u_t | u_{t-1}, z_{t-1}, u_{t-2}, z_{t-2} \ldots) = E(u_t | u_{t-1}, u_{t-2}, \ldots). \tag{6.1}$$

Assuming linear expectations and first order dynamics, this is the same as

$$E(u_t | u_{t-1}, z_{t-1}, u_{t-2}, z_{t-2} \ldots) = \rho u_{t-1}. \tag{6.2}$$

Because the static/AR(1) model implies that

$$E(u_t | z_t, u_{t-1}, z_{t-1}, u_{t-2}, z_{t-2} \ldots) = \rho u_{t-1}, \tag{6.3}$$

(6.2) is generally weaker than the assumptions underlying the static/AR(1) model, unless $\{z_t\}$ is assumed to be strictly exogenous. Examining the implications of and how to test conditions like (6.1) deserves further research, but is beyond the scope of the current paper.

References

Berk, K.N. (1974), "Consistent Autoregressive Spectral Estimates," *Annals of Statistics* 2, 489-502.

Domowitz, I. and H. White (1982), "Maximum Likelihood Estimation of Misspecified Models," *Journal of Econometrics* 20, 35-58.

Engle, R.F. (1982), "Autoregression Conditional Heteroskedasticity with Estimates of United Kingdom Inflation," *Econometrica* 50, 987-1008.

Engle, R.F. (1984), "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*, Vol. II. Amsterdam: North Holland, 775-826.

Gallant, A.R. (1988), *Nonlinear Statistical Models*. New York: Wiley.

Gallant, A.R. and H. White (1988), *A Unified Approach to Estimation and Inference in Nonlinear Dynamic Models*. Oxford: Basil Blackwell.

Godfrey, L.G. (1987), "Discriminating Between Autocorrelation and Misspecification in Regression Analysis: An Alternative Test Strategy," *Review of Economics and Statistics* 69, 128-134.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

Hansen, L.P. and R.J. Hodrick (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy* 88, 829-853.

Harvey, A.C. (1981), *The Econometric Analysis of Time Series*. Oxford: Philip Alan.

Hausman, J.A. (1978), "Specification Tests in Econometrics" *Econometrica* 43, 727-738.

Hausman, J.A. (1983), "Specification and Estimation of Simultaneous Equation Models," in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*, Vol. I. Amsterdam: North Holland, 391-448.

Hendry, D.F. and G.F. Mizon (1978), "Serial Correlation as a Convenient Simplification, not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England," *The Economic Journal* 88, 549-563.

Hendry, D.F., A.R. Pagan, and J.D. Sargan (1984), "Dynamic Specification," in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*, Vol. II. Amsterdam: North Holland, 1023-1100.

Hsieh, D.A. (1983), "A Heteroskedasticity-Consistent Covariance Matrix Estimator for Time Series Regressions," *Journal of Econometrics* 22, 281-290.

Newey, W.K. and K.D. West (1987), "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

Pagan, A.R. and A.D. Hall (1983), "Diagnostic Tests as Residual Analysis," *Econometric Reviews* 2, 159-218.

Thursby, J.G. (1985), "A Test Strategy for Discriminating Between Autocorrelation and Misspecification in Regression Analysis: A Reply," *Review of Economics and Statistics* 67, 177-178.

White, H. (1980), "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48, 817-838.

White, H. (1984), *Asymptotic Theory for Econometricians.* Orlando: Academic Press.

White, H. and I. Domowitz (1984), "Nonlinear Regression with Dependent Observations," *Econometrica* 52, 143-162.

Wooldridge, J.M. (1989), "A Computationally Simple Heteroskedasticity and Serial Correlation Robust Standard Error for the Linear Regression Model," *Economics Letters* 31, 239-243.

Wooldridge, J.M. (1990a), "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory* 6, 17-43.

Wooldridge, J.M. (1990b), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," forthcoming, *Journal of Econometrics* issue on "Diagnostic Testing."